

The emergence of classifiers from class term compounds in Vietnamese

Mike Pham

University of Chicago

SEALS23, 2013

The big questions

- Where do (new) classifiers come from?
- Why do classifiers have the semantic restrictions they do?
- What are the conditioning factors for the emergence of classifiers?

Numeral classifiers

- Classifiers are well-known to be used in (South)East Asian languages to count nouns:

(1) a. bốn *(**con**) chó
four CLF:ANI dog
'four dogs'

b. mười *(**trái**) cam
ten CLF:FRUIT orange
'ten oranges'

- Classifiers organize nouns based on general properties (i.e. animacy, shape, function, etc):

Lexical sources of classifiers

- Classifiers most often come from nouns (Aikhenvald, 2000)

- (2) a. Hai loại **lá**
two CLF:KIND leaf
'two kinds of leaves'
- b. Hai **lá** thư
two CLF:SHEET letter
'two letters'

Lexical sources of classifiers

- Classifiers most often come from nouns (Aikhenvald, 2000)

(2) a. Hai loại **lá** b. Hai **lá** thư
two CLF:KIND leaf two CLF:SHEET letter
'two kinds of leaves' 'two letters'

- The semantic restriction of classifiers seems related to the most salient properties of the source noun

Lexical sources of classifiers

- Classifiers most often come from nouns (Aikhenvald, 2000)

(2) a. Hai loại **lá** b. Hai **lá** thư
two CLF:KIND leaf two CLF:SHEET letter
'two kinds of leaves' 'two letters'

- The semantic restriction of classifiers seems related to the most salient properties of the source noun
- Simplistically, we can say $N > Clf$; but what is the intermediate stage in this process?

Lexical sources of classifiers

- Classifiers most often come from nouns (Aikhenvald, 2000)

(2) a. Hai loại **lá** b. Hai **lá** thư
two CLF:KIND leaf two CLF:SHEET letter
'two kinds of leaves' 'two letters'

- The semantic restriction of classifiers seems related to the most salient properties of the source noun
- Simplistically, we can say $N > Clf$; but what is the intermediate stage in this process?
- $N > ? > Clf$

Class term compounds

- Class terms do not categorise the element to which they attach, but rather the whole compound of which they are part (Enfield, 2004)
- That is, class terms have a **taxonomic** relation to the compounds they derive

(3) a. [má y chũ]
machine grapheme
'typewriter'

b. [má y bay]
machine fly
'airplane'

c. [má y giặ t]
machine wash
'washing machine'

d. [má y lạn h]
machine cold
'air conditioner'

Class terms as medial stage

- "The difference between derivational morphemes and class terms that are used in compounding processes is not always worked out, and both have been labelled 'classifiers' at times" (Grinevald, 2000, p.59)

Class terms as medial stage

- "The difference between derivational morphemes and class terms that are used in compounding processes is not always worked out, and both have been labelled 'classifiers' at times" (Grinevald, 2000, p.59)
- Vietnamese class term compounds show mixed properties of bare NPs and ClfPs

Class terms as medial stage

- "The difference between derivational morphemes and class terms that are used in compounding processes is not always worked out, and both have been labelled 'classifiers' at times" (Grinevald, 2000, p.59)
- Vietnamese class term compounds show mixed properties of bare NPs and ClfPs
- Class terms are a major source for the development of new classifiers in Thai (and Proto-Tai) (DeLancey, 1986)

Class terms as medial stage

- "The difference between derivational morphemes and class terms that are used in compounding processes is not always worked out, and both have been labelled 'classifiers' at times" (Grinevald, 2000, p.59)
- Vietnamese class term compounds show mixed properties of bare NPs and ClfPs
- Class terms are a major source for the development of new classifiers in Thai (and Proto-Tai) (DeLancey, 1986)
- noun > class term > classifier

Proposal:

- Following DeLancey (1986), class term compounding is productive in **grammaticalization of nouns to classifiers** in Vietnamese as well
- Class term compounding involves **semantic bleaching**; the remaining meaning corresponds to the semantic restrictions characteristic of classifiers
- Class terms are **reanalyzed** as classifiers in Vietnamese due to surface similarity between class term compounds and ClfPs
 - ▶ head-modifier word order
 - ▶ disyllabic prosodic preference

The Roadmap

- 1 Class term compounds
 - NP-like behavior
 - Classifier-like behavior
- 2 New classifier emergence
 - Semantic bleaching
 - Extension to non-compound nouns
 - Traces of emergence
- 3 Reanalysis
 - Head-modifier relation
 - Disyllabic prosodic preference

Outline

- 1 Class term compounds
 - NP-like behavior
 - Classifier-like behavior
- 2 New classifier emergence
 - Semantic bleaching
 - Extension to non-compound nouns
 - Traces of emergence
- 3 Reanalysis
 - Head-modifier relation
 - Disyllabic prosodic preference

Mixed properties of class term compounds

- Class term compounds behave like both bare NPs and ClfPs:

NP-like	ClfP-like
generic interpretation underspecified for number can't be coordinated	(directly countable) (NP-Ellipsis)

Semantic interpretations

- Kirby (2006) shows that classifiers restrict semantic meaning:

	Bare NP	[Clf+N]	[Num+Clf+N]
Definiteness	Indef/Def/Gen	Indef/Def	Indef
Number	Sg/Pl	Sg	—

- Bare nouns allow generic interpretations and are underspecified for number
- Classifier phrases cannot have generic interpretations and are singular (without a numeral)

Generic interpretations

- Bare nouns allow generic interpretations:

(4) **Chó** ăn thít.
dog eat meat
'Dogs eat meat.'

- Compounds also allow generic interpretations:

(5) **Máy bay** đi lẹ lắm.
machine fly go quick very
'Airplanes go really quickly.'

Number ambiguity

- Bare nouns are underspecified for number:

(6) Tôi có **sách** ở nhà.

1.SG have book at house

'I have (a/the) book(s) at home.'

- Compounds are also underspecified for number:

(7) Tôi có **máy tính** ở nhà.

1.SG have machine calculate at home

'I have (a/the) computer(s) at home.'

Coordination

- It is possible to coordinate nouns under a shared classifier:

(8) một trăm **trái** chuối với cam
one hundred CLF:FRUIT banana and orange
'one hundred [oranges and bananas]'

- If class terms are actually classifiers we would expect coordination of their complements. This is not the case:

(9) *một trăm **máy** bay với giặt
one hundred machine fly and wash
'intended: one hundred [airplanes and washing machines]'

Classifier-like behavior: Counting nouns

- Recall that nouns in Vietnamese generally require classifiers in order to be counted:

(1a) hai *(con) chó
two CLF:ANI dog
'two dogs'

(1b) bốn *(trái) cam
four CLF:FRUIT orange
'four oranges'

Counting compounds

- Many class term compounds can be directly counted, where the class term appears to be acting as a classifier

- (10) a. bốn **nhà thương**
four house injure
'four hospitals'
- b. hai **phòng khách**
two room guest
'two living rooms'
- (11) a. tám (chiếc) **máy bay**
eight CLF:UNIT machine fly
'eight airplanes'
- b. năm (người) **thợ may**
five CLF:PERSON worker sew
'five tailors'

Counting compounds

- Many class term compounds can be directly counted, where the class term appears to be acting as a classifier

(10) a. bốn **nhà thương**
four house injure
'four hospitals'

b. hai **phòng khách**
two room guest
'two living rooms'

(11) a. tám (chiếc) **máy bay**
eight CLF:UNIT machine fly
'eight airplanes'

b. năm (người) **thợ may**
five CLF:PERSON worker sew
'five tailors'

- Because classifiers can't stack, optional classifiers show that class terms cannot be full classifiers

NP-Ellipsis (NPE)

- Classifiers license NPE, where the NP complement can be elided if it is contextually retrievable (i.e. linguistic antecedent) (Nguyễn, 2004; Alexiadou and Gengel, 2011; Cheng and Sybesma, 2009; Jenks, 2011):

(12) Nếu mày đọc ba **cuốn** sách, tôi sẽ đọc ba
if 2.SG read three CLF:VOLUME book 1.SG FUT read three
cuốn luôn.
CLF:VOLUME also
'If you read three books, I will read three books as well.'

"NP"-Ellipsis

- Class terms also seem to license ellipsis of their complements:

(13) Nếu màỵ mua ba **máy lạnh**, tôi sẽ mua ba
if 2.SG buy three machine cold 1.SG FUT buy three
máy luôn.
machine also
'If you buy three air conditioners, I will buy three ones as well.'

Interim Summary:

- Class terms derive a taxonomy of compounds, which show properties of both classifier phrases and bare nouns:

NP-like	ClfP-like
generic interpretation underspecified for number can't be coordinated	(directly countable) (NPE)

- Class term compounds are NPs that have taken on ClfP properties
- These mixed characteristics are expected if $N > \text{class term} > \text{Clf}$

Outline

- 1 Class term compounds
 - NP-like behavior
 - Classifier-like behavior
- 2 New classifier emergence
 - Semantic bleaching
 - Extension to non-compound nouns
 - Traces of emergence
- 3 Reanalysis
 - Head-modifier relation
 - Disyllabic prosodic preference

Expanding the semantic field

- Why do classifiers have the semantic criteria they do?

Expanding the semantic field

- Why do classifiers have the semantic criteria they do?
- Semantic bleaching in class term compounding provides an answer

Expanding the semantic field

- Why do classifiers have the semantic criteria they do?
- Semantic bleaching in class term compounding provides an answer
- Compounds can have idiomatic meanings:
- **máy lạnh**, lit. ‘machine cold’, does not denote a cold machine, but specifically an air conditioner

Expanding the semantic field

- Why do classifiers have the semantic criteria they do?
- Semantic bleaching in class term compounding provides an answer
- Compounds can have idiomatic meanings:
- **máy lạnh**, lit. ‘machine cold’, does not denote a cold machine, but specifically an air conditioner
 - ▶ Also note that this does not denote other cold-making machines, such as refrigerators

Bleaching the class term

- Because a class term taxonomizes its compounds, its semantic field must be expanded to include these new meanings

Bleaching the class term

- Because a class term taxonomizes its compounds, its semantic field must be expanded to include these new meanings
- Its noun source has some specific lexical meaning...

Bleaching the class term

- Because a class term taxonomizes its compounds, its semantic field must be expanded to include these new meanings
- Its noun source has some specific lexical meaning...
- However, the meaning of the class term becomes less specific during compounding in order to generalize across the growing semantic field

Generalized meanings

- The unbleached meanings used for compounding correlate to the most salient properties of the noun:
 - ▶ Synchronically, *xe* means 'car', but as a class term means something closer to 'vehicle':

(14) a. *xe đạp*
vehicle step
'bicycle'

b. *xe lửa*
vehicle fire
'train'

- This could explain why classifiers tend to organize nouns by physical features, function and social interaction (Denny, 1976)

Defense of bleaching

- Why not assume that *xé* just means ‘vehicle’ from the start?
- Consider examples where there is more than one salient property

(15) **bánh**, ‘cake’

- a. **bánh chưng**, ‘rice dumpling’ (lit. ‘cake stew’)
 - b. **bánh bao**, ‘steamed bun’ (lit. ‘cake bag’)
 - c. **bánh mì**, ‘bread’ (lit. ‘cake wheat’)
 - d. **bánh phở**, ‘rice noodles’ (lit. ‘cake phở’)
 - e. **bánh xe**, ‘wheel’ (lit. ‘cake vehicle’)
 - f. **bánh răng**, ‘cog’ (lit. ‘cake teeth’)
- Some *bánh* compounds pick out food properties, while others pick out shape properties

- This divergence in salient properties also appears in classifiers, which is consistent with an analysis where they are derived from class terms

- (16) a. mười **trái** cam
ten CLF orange
'ten oranges'
- b. mười **trái** chuối
ten CLF banana
'ten bananas'
- c. mười **trái** banh
ten CLF ball
'ten balls'

Extension to non-compound nouns

- Recall that *xe* means 'car', and can derive compounds denoting vehicles:

(17) a. *xe đạp*
vehicle step
'bicycle'

b. *xe lửa*
vehicle fire
'train'

- To fully emerge as a classifier, a class term must be used in a context where it can be reanalyzed as a classifier
- These will be contexts where the meaning the class term contributes is somewhat redundant

- *xe* is also used as a class term for loanwords denoting wheeled vehicles
- Loanwords should already have idiomatic meaning without the need for compounding; *xe* seems redundant

(18) a. ba **xe** tácx
 three XE taxi
 ‘three taxis’

b. tám **xe** táng
 eight XE tank
 ‘eight tanks’

- Xe seems to be emerging as a full classifier for wheeled vehicles

(19) Melia Hà Nội có 2 xe Mercedes [sic] 16 chỗ, 1 xe
 Melia Hanoi have two XE Mercedes sixteen seat, one XE
 Limúuine [sic] Mercedes Benz [sic] và một đội gồm 10 xe
 limousine Mercedes Benz and one team comprise ten XE
 tắc xi
 taxi

‘Melia Hanoi has two sixteen-seat Mercedes [vans], one Mercedes Benz limousine and a team comprising ten taxi cabs.’

<http://www.doko.vn/luan-van/Bao-cao-thuc-tap-115351> (November 30, 2012)

From trees to long skinny things

- If this emergence story is true, we might expect to see certain words as nouns, class terms and classifiers; i.e. traces of classifier emergence from a noun via compounding

- (20) a. cây, 'tree(s)/plant(s)' noun
- b. cây đàn
tree play.music
'musical instrument(s)' class term
- c. cây bút
CLF:LONG-SKINNY pen
'a/the pen' classifier

Summary

- Nouns can be used as class terms in compounding, where they are bleached of non-salient meanings
- The semantic properties that are not bleached become the semantic restriction of classifiers
- Compounds involving loanwords are cases in which class terms are reanalyzed as classifier

Outline

- 1 Class term compounds
 - NP-like behavior
 - Classifier-like behavior
- 2 New classifier emergence
 - Semantic bleaching
 - Extension to non-compound nouns
 - Traces of emergence
- 3 Reanalysis
 - Head-modifier relation
 - Disyllabic prosodic preference

Reanalysis is driven by two primary factors:

- The semantic head-modifier relation of class term compounds and ClfP
- Disyllabic prosodic preference in Vietnamese

Compound word order

- Class term compounds are head-initial, with no categorical restrictions on the modifying element: [CT X]

Compound word order

- Class term compounds are head-initial, with no categorical restrictions on the modifying element: [CT X]
- The modifier specifies the meaning of the class term:

Compound word order

- Class term compounds are head-initial, with no categorical restrictions on the modifying element: [CT X]
- The modifier specifies the meaning of the class term:
- It is a function that takes the set denoted by the class term and returns a specific subset of that

Compound word order

- Class term compounds are head-initial, with no categorical restrictions on the modifying element: [CT X]
- The modifier specifies the meaning of the class term:
- It is a function that takes the set denoted by the class term and returns a specific subset of that
 - ▶ In *xe đạp*, (lit. ‘vehicle step’), *đạp* is a function that takes the set of vehicles and returns the (sub)set of bicycles (step-vehicles)

ClfP word order

- ClfPs are head-initial, with a categorical restriction on their modifiers being nouns: [Clf NP]

ClfP word order

- ClfPs are head-initial, with a categorical restriction on their modifiers being nouns: [Clf NP]
- Classifiers denote individual atoms with general qualities: animacy, shape, function, etc.

ClfP word order

- ClfPs are head-initial, with a categorical restriction on their modifiers being nouns: [Clf NP]
- Classifiers denote individual atoms with general qualities: animacy, shape, function, etc.
- The modifying NP specifies what kind of atoms these are:

ClfP word order

- ClfPs are head-initial, with a categorical restriction on their modifiers being nouns: [Clf NP]
- Classifiers denote individual atoms with general qualities: animacy, shape, function, etc.
- The modifying NP specifies what kind of atoms these are:
- It is a function that takes the set (of atoms) denoted by the classifier and returns a specific subset (of atoms) of that

ClfP word order

- ClfPs are head-initial, with a categorical restriction on their modifiers being nouns: [Clf NP]
- Classifiers denote individual atoms with general qualities: animacy, shape, function, etc.
- The modifying NP specifies what kind of atoms these are:
- It is a function that takes the set (of atoms) denoted by the classifier and returns a specific subset (of atoms) of that
 - ▶ In *trái cam*, (lit. ‘fruit orange’), *cam* is a function that takes the set of individual fruit and returns the (sub)set of individual oranges (orange-fruit)

Semantic relation-driven reanalysis

- Class term compounds and ClfPs both have the same subjective head-modifier relation

Semantic relation-driven reanalysis

- Class term compounds and ClfPs both have the same subsecutive head-modifier relation
- Because *xe đạp*, 'bicycle', is a lexicalized compound, the subset returned is idiomatic/unpredictable

Semantic relation-driven reanalysis

- Class term compounds and ClfPs both have the same subjective head-modifier relation
- Because *xe đạp*, ‘bicycle’, is a lexicalized compound, the subset returned is idiomatic/unpredictable
- But in cases like *xe tấng*, ‘tank’, the subset returned is predictable, as they are for ClfPs

- ClfPs are special cases morphosyntactically, as they only allow NP modifiers
- Reanalysis will then first occur in compounds that have a [CT NP] structure
- Classifiers also have the individuation function, which is then composed into the semantics of the class term during reanalysis (von Stechow, 1995)
- Eventually, the class term morpheme is base-generated in the Clf position, marking full emergence

Disyllabic prosodic preference

- Thompson (1965) briefly mentions a general preference for disyllabic structures in Vietnamese (such as for fragment answers and secondary predicates)
- Tran (2011) looked at child acquisition of Vietnamese classifiers, and found that children often failed to produce the correct classifier when presented with disyllabic nouns
 - ▶ These included loanwords, (class term) compounds, reduplicated nouns
 - ▶ Two strategies children employed were omitting the classifier, and including the classifier but making the noun monosyllabic

- There seems to be a preference for disyllabic structures in Vietnamese generally speaking, but especially when following a numeral
- This prosodic pressure, alongside the semantic head-modifier relation, also drives reanalysis in class term compounds

Conclusion

Take-home points

- Class term compounding productively grammaticalizes nouns into classifiers: N > CT > Clf
- Class terms which are highly productive in compounding are likely to be semantically bleached, retaining only general semantic properties corresponding to a classifier's semantic restrictions on its NP complement
- In Vietnamese, reanalysis of a class term as a classifier is motivated by the surface word order of the head-modifier semantic relation, as well as prosodic preference for disyllabic structures following numerals

Further research

- Overhauling classifier semantics to reflect my claim that classifiers don't individuate the denotation of their NP complement, but the denotation of the entire [Clf NP] structure
- How to predict speaker variation in classifier omission?
- How do prosody and syntax-semantics interact?
- Synchronic analysis as well?

Thank you!
Cám ơn!

Acknowledgements: Thanks to Karlos Arregi, Itamar Francez and Jason Riggle for their linguistic intuitions; and to Huonglien Pham, Chieu Nguyen, Karyn Mai, and other survey-takers for their Vietnamese intuitions. All mistakes are my own.

References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford studies in typology and linguistic theory. New York: Oxford University Press Inc.
- Alexiadou, Artemis, and Kirsten Gengel. 2011. Classifiers as morphosyntactic licensors of NP ellipsis: English vs. Romance. In *Proceedings of NELS 39*, 15–28. Manuscript. Universität Stuttgart.
- Cheng, Lisa Lai-Shen, and Rint Sybesma. 2009. De as an underspecified classifier: first explorations. *Yǔyánxué lùncóng* 39.
- DeLancey, Scott. 1986. Toward a history of Tai classifier systems. In *Noun Classes and Categorization*, ed. Colette Craig, volume 7 of *Typological studies in language*, 437 – 452. Amsterdam/Philadelphia: John Benjamins B.V.
- Denny, J. P. 1976. What are noun classifiers good for? In *Proceedings from the twelfth regional meeting of the Chicago Linguistics Society*, 122–123.
- Enfield, Nick J. 2004. Nominal classification in Lao: a sketch. *Sprachtypologie und Universalienforschung* 57:117–143.
- von Fintel, Kai. 1995. The formal semantics of grammaticalization. In *NELS 25 Proceedings*, volume 2: Papers from the Workshops on Language Acquisition and Change, 175–189. GLSA, UMass Amherst.
- Grinevald, Colette. 2000. *Systems of nominal classification*, chapter A morphosyntactic typology of classifiers, 50–92. Number 4 in *Language, culture and cognition*. Cambridge, UK: Cambridge University Press.
- Jenks, Peter Spiros Eric. 2011. The hidden structure of Thai noun phrases. Doctoral Dissertation, Harvard University, Cambridge, Massachusetts.
- Kirby, James. 2006. Vietnamese and the structure of NP. U. Chicago manuscript.
- Nguyễn, Tuong Hung. 2004. The structure of the Vietnamese noun phrase. Doctoral Dissertation, Boston University.
- Thompson, Laurence C. 1965. *A Vietnamese grammar*. Number XIII-XIV in *Mon-Khmer Studies*. University of Hawai'i Press.
- Tran, Jennie. 2011. The acquisition of Vietnamese classifiers. Doctoral Dissertation, University of Hawai'i at Manoa.

Survey data

- 100 fill in the blank questions:
 - ▶ 23 head nouns
 - ▶ 77 compounds derived from those nouns (pulled from dictionary)
 - ▶ 10 fillers
- 32 survey takers

a. Overall totals (collapsed)

totals	Y	N
Y	1125	489
N	51	301

b. Overall percentages (collapsed)

totals	Y	N
Y	57	25
N	3	15